

基于 k 匿名的混合式隐私保护系统的设计

杜永文, 朱晓阳, 练云翔

(兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070)

摘 要: 基于位置服务提出了基于 k 匿名的混合式隐私保护系统(HPPS), 分别阐述了混合式隐私保护系统中分散式和中心式两部分的设计. 实验结果表明, 当 k 值增大时, 隐私保护程度也在增加, 同时兴趣点查询质量和隐私保护之间也达到了一种平衡.

关键词: 基于位置服务; 隐私保护架构; 兴趣点; k 匿名; 查询质量

中图分类号: TP309

文献标识码: A

文章编号: 1000-7180(2016)03-0056-04

Design of Hybrid Privacy Protection System Based on k Anonymous

DU Yong-wen, ZHU Xiao-yang, LIAN Yun-xiang

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: This paper proposes hybrid privacy protection system (HPPS) based on k anonymous, and it expounds the design of the two part of the system that contain the decentralized and centralized architecture. Experimental results show that the level of privacy protection is increasing while the k value increases. At the same time, the query quality of POI and privacy protection has reached a balance.

Key words: location based services (LBS); framework of privacy protection; point of interest (POI); k anonymous; query quality

1 引言

移动用户在 LBS 服务中向位置服务提供商请求兴趣点信息, 这种用户发送当前位置信息给服务提供商的模式, 不可避免地造成位置信息的泄露. 随着用户对位置敏感信息的关注, 如何保护移动网络中的位置信息成为隐私保护领域的研究热点.

移动网络中在不同架构下已有研究者提出不同的隐私保护方法. 在第三方可信服务器架构下, 综合 k 匿名和区域化位置信息两种思想, 提出了一种能够处理大量并发用户的位置隐私保护方法^[1]. 还有研究将模糊化 (Obfuscation) 作为位置隐私保护的一种有效机制, 利用位置模糊化区域来进行模糊查询处理, 仍然可以保持服务质量的一种隐私保护算法^[2]. Silent Cascade 算法从空间和时间这两个方面对位置信息进行处理^[3]. 上述隐私保护架构主要分

为两种, 一种是分散式隐私保护架构^[4], 另一种是中心式隐私保护架构^[5]. 但是分散式结构存在以下问题: (1) 移动客户端需要足够大的存储空间和计算能力; (2) 无法保证系统的可靠性^[6]. 中心式结构存在以下问题: (1) 中心服务器的可信度; (2) 中心服务器的负载能力^[7]. 基于此, 本文提出了介于分散式和中心式之间并且结合 k 匿名技术的混合式隐私保护系统 HPPS (Hybrid Privacy Protect System), 即采用分散式和中心式相结合的方式, 通过分别在每个区域的中心服务器设置特殊的处理程序, 实现客户端与客户端以及客户端与中心服务器的通信, 以达到保护用户真实位置信息的目的.

2 HPPS 架构设计

HPPS 的架构设计主要分为三大部分, 包括移动终端、中心匿名服务器和 LBS 服务器, 在同一个

区域内移动终端可以有多个,并且在一个区域内的中心匿名服务器仅有一台,而 LBS 服务器可以被不同区域内的中心匿名服务器所公用.图 1 为混合式隐私保护系统的架构图,给出了某一个区域内的这三部分的模型.

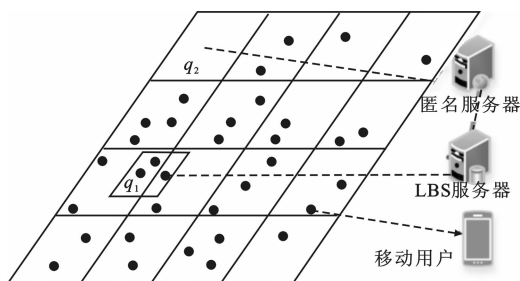


图 1 HPPS 架构图

本文的 HPPS 架构中结合了 Sweeney 提出的 k 匿名方法^[8]. 所谓 k 匿名是查询者为了不要暴露自己的位置, 利用某些方法把自己的位置与其他 $k-1$ 个使用者位置做成隐藏的处理方式. 为了达到隐藏的目的又要保持地址位置查询内容的正确, 查询者会与邻近的使用者计算出一块包含 k 需求的匿名区域来代替自己的位置. 这样匿名的区域会是包含使用者而不是过于庞大的区域, 此时, 查询者可将查询内容与隐私区域当作查询的条件做兴趣点查询.

从图 1 中的 q_1 可看出, 移动用户发送查询信息时, 移动设备会先按使用者所需的隐私保护程度 k 值以分散式网络方式连接周围的使用者. 当使用者和社交网络中周围的使用者共同组成符合 k 匿名的条件时, 即可采用分散式隐私方式组成模糊区域^[9]. 但是, 当使用者与周围连接数不足时, q_2 移动设备会自动切换到中心匿名服务器的隐私保护方式.

混合式隐私保护系统 HPPS 是兴趣点查询的请求节点,分散式网络中的其他用户也为协助节点和当前区域的中心匿名服务器以及 LBS 服务器共同组成,在请求节点与协助节点共同满足 k 的情况下使用分散式部分,不满足 k 的情况下请求中心匿名服务器系统,经过中心匿名服务器的处理之后,查询请求与隐私区域一并传送给 LBS 服务器做查询请求,当返回了 LBS 服务器的查询请求之后,匿名服务器会进行筛选处理,直到把兴趣点查询返回给用户请求节点为止,它们之间的时序交互方式如图 2 所示。

3 HPFS 分散式部分设计

分散式隐私保护方式即查询者以分散式网络的

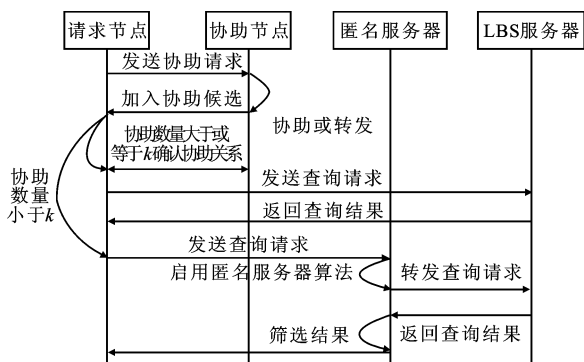


图 2 HPPS 交互时序图

连接方式以最多 2 个 hop 与周围的使用者进行连接,因而可以构成连续性的隐私保护区域,得到周围可以连接的使用者与其相对速度和方向,而使用者会根据周边使用者数量和选择算法选择其中一个设备当作代理者送出连续性的隐私区域给 LBS 服务器. LBS 会依照此区域搜索兴趣点信息,分散式部分伪代码如下所示.

输入:兴趣点查询信息

输出:分散式模糊化区域

参数:与周围移动用户的连接跳数 hop

(1) for(查询者附近的每个有效的连接的数量) then

(2) 查询矩形区域内的所有近邻者

(3) 找出相关的四个位置点

(4) end for

(5) for(最大延迟时间) then

(6) 查询附近的用户数量

(7) 确定模糊化区域

(8) 寻找一位代理者发送模糊化区域给 LBS 服务器

(9) end for

4 HPPS 中心式部分设计

中心匿名服务器部分是由查询者所属的匿名服务器与周围区域内的匿名服务器所负责,当使用者无法由分散式方式找到适当的隐私区域时,会将查询请求发送给自身所在区域的中心匿名服务器,而匿名服务器就会依照使用者的位置速度和移动方向来寻找查询者所需要的隐私需求 k ,中心服务器会先搜索本区域内的使用者个数是否满足查询者的匿名需求,若是不够时,中心匿名主机会寻求周边中心匿名主机的协助。

关于 HPPS 中心匿名服务器部分首先使用 IF 来判断用户所在区域的用户数量是否满足用户兴趣点查询请求所制定的 k 值,当中心匿名服务器的用户数量不满足 k 值请求时,我们对中心区域的扩展进

行了设计,采用了 while 循环的方式,直到找出满足 k 值的匿名空间为止,中心式部分伪代码如下所示.

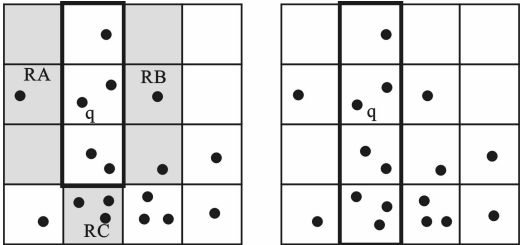
输入:兴趣点查询信息

输出:中心匿名服务器模糊化区域、匿名区域 k 值、用户所在区域中心坐标 $user[x][y]$

参数:用户数量、区域总数

```
(1)  if(用户  $user[x][y]$  所在区域的用户数量  $> k$ ) then
(2)      return 用户  $user[x][y]$  所在区域为隐私区域
(3)  else
(4)      while( 总的用户数量  $< k$  )
(5)          分别计算周边东南西北四个方向区域内的用户数量
(6)          找出用户数量最多的一个区域
(7)          if( 其中两个区域内数量相同 )
(8)              分别计算区域外围有活跃用户的区域总数
(9)              选择区域总数较大的区域为扩展方向
(10)         end if
(11)         确定模糊化区域
(12)         寻找一位代理者发送模糊化区域给 LBS 服务器
(13)     end while
(14) end if
```

隐私区域成长方式为依照最多隐私数的地方扩展.如图 3(a)所示,因为区域 RC 的隐私使用者最多,所以以 RC 作为扩展,这样的隐私区域可为长方形的隐私区域图,如图 3(b)所示,以下对这样的长方形隐私区域与原本的隐私区域的查询时间与难度做比较.



(a) 原始隐私区域 (b) 扩展后隐私区域

图 3 隐私区域扩展示意图

假设我们隐私区域每个单位边有 n 个需要判断是否为查询结果的兴趣点,矩形的变得边总长为 p ,以最极端的矩形四边来做例子,1,1,($p-2$)/2,($p-2$)/2 分别为矩形的四边,而计算 k 邻近点查询所需要的时间复杂度总和推导如下:

$$O(1 + 1 + 1 + 1) + 2 * ((\frac{p-2}{2} * n) \log(\frac{p-2}{2} * n)) \Rightarrow O((p-2) * n \log \frac{(p-2) * n}{2}) \Rightarrow O(pn \log \frac{pn}{2})$$

由此可以看出原本的 HPPS 做法的时间复杂度为 $O((pn) \log(\frac{pn}{4}))$,而不限定长宽差距的做法的

时间复杂度为 $O((pn) \log(\frac{pn}{2}))$,在长宽度差越大的矩形会影响其搜索所得的排序时间.在兴趣点的均匀分布的空间下的环境下,线上 k 邻近点搜索的时间难度会因为线的长度而变化,线越长代表我们要排序的兴趣点越多,因此采用长宽距离对等的矩形可以减少 LBS 信息搜索的运算时间.

5 实验结果与分析

Random-way-point^[10]是移动网络中一种常用的随机模型,利用此模型对 HPPS 进行模拟实验.实验以 5 000 个移动查询者随机生成 1 000 个查询内容,而查询的兴趣点有 10 000 个随机分布在空间中,而查询内容均为最近的 10 个邻近点的查询,每个查询者都可以与 2-hop 内的使用者做分散式隐私保护的请求.在此实验环境中,把隐私保护空间分割成 32 个大小相同的子空间,只要分散式方法失败马上会启动中心匿名服务器方式来做隐私查询保护,以中心服务器管理范围内的所有使用者位置以及速度向量.每个使用者的速度有 0 ~ 1 000 m/min 随机在 16 000 m * 16 000 m 的环境中,详细的实验设计如表 1 所示.

表 1 实验参数设定

参数	参数说明	参数选取范围
空间面积	总的空间范围	16 000 m * 16 000 m
分割面积	每个分割区域的大小	500 m * 500 m
用户数量	分布在空间内的用户总数	50 000
兴趣点数量	分布在空间内的兴趣点总数	10 000
查询次数	用户的兴趣点查询次数	1 000
分散式最大跳步数	允许信息传播的条步数	2 - hop
k 值	允许 k -匿名的匿名度量	1 ~ 100
P 值	p -NN 查询参数	10
移动速度/(m/min)	用户移动速度	0 ~ 1 000

由图 4 可知,HPPS 系统的分散式以及中心式隐私的使用率在不同隐私需求下的使用比例,在隐私参数 $k \leq 25$ 时的状况下,HPPS 的分散式隐私方式都可以正常地处理隐私需求并且可以正常的返回隐私查询的结果.当 k 值的需求较低时,分散式的使用率较高,随着 k 值的增长,周边的隐私参与节点的数量不能满足请求节点的隐私需求,转而使用中心匿名服务器,因此 k 值的增大使得中心匿名服务器参与进来生成更大范围的匿名空间以满足用户的隐私

要求。

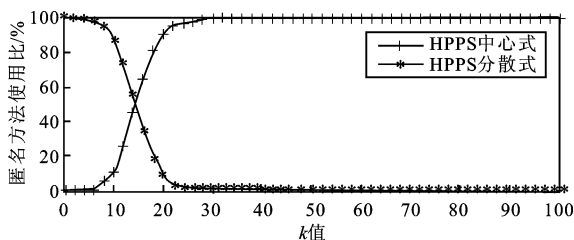


图4 HPPS 各部分隐私方法使用率趋势图

图5表示查询者所需最低隐私需求以及实际隐私系统完成隐私的实际状况,隐私数的大小代表隐私保护的安全程度,但安全性越高,会带来的系统效率的降低,所以一个好的隐私保护系统必须有好的隐私状态并且能够保持高效率。随着 k 值的增大隐私数呈上升区域,且当 $k \geq 70$ 时,隐私数趋于稳定。

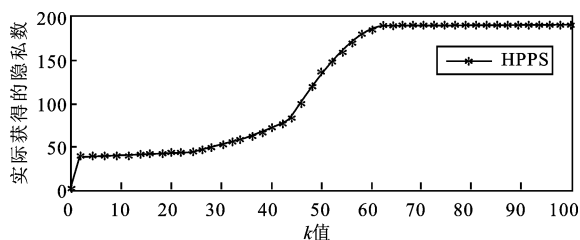


图5 实际获得的隐私数随隐私参数 k 变化图

由图6可知,查询者随着 k 值的变化所返回的总的兴趣点候选集合趋势图,查询者位置在隐私保护系统的位置模糊下获得隐私的相对安全而对LBS服务来说,使用者的位置是一个区域,所以按照查询内容获得的查询结果来说也会因此多包括一些防止误差过大的结果集合,用户查询兴趣点时,想要得到真实的查询答案时必须把等待经过答案的筛选,执行过滤才能获得。然而隐私保护的方式会影响待选答案的数量以及结果的正确性,结果的正确性以及查询动作的效率是用户在兴趣点查询时最在乎的一个方面。所以在本文的试验中,隐私需求 k 在25左右以下时,HPPS方法均由分散式部分进行隐私处理。在 k 大于25之后,图中表现出来的HPPS的待选答案数量成长较为平缓。由此可以得知,稳定的隐私区域成长会对查询的质量有很大的帮助。

6 结束语

本文设计了一种基于 k 匿名的混合式隐私保护系统,该位置隐私保护系统结合了分散式和中心式架构,同时避免了分散式架构无法保障可靠性和中心式架构的瓶颈问题。该系统结合 k 匿名模型通过用户指定 k

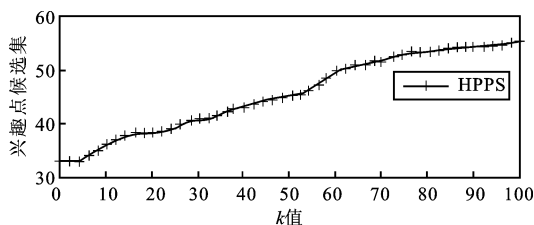


图6 兴趣点候选集随隐私参数 k 的变化趋势图

值来确定隐私保护程度。实验结果表明,随着 k 值的不断变化,从实际得到的隐私数和兴趣点候选集的变化趋势图中可分析出最佳的隐私保护 k 值,在隐私保护的同时也保证了兴趣点信息的查询质量。

参考文献:

- [1] 赵泽茂,胡慧东,张帆,等. 圆形区域划分的 k -匿名位置隐私保护方法[J]. 北京交通大学学报, 2013, 37(5):13-18.
- [2] Duckham M, Kulik L. A formal model of obfuscation and negotiation for location privacy[C] //PERVASIVE 2005. Germany, Munich, 2005:152--170.
- [3] 王彩梅,郭亚军,郭艳华. 位置服务中用户轨迹的隐私度量[J]. 软件学报, 2012, 23(2):352-360.
- [4] 朱青,赵桐,王珊. 面向查询服务的数据隐私保护算法[J]. 计算机学报, 2010, 33(8):1315-1323.
- [5] 赵耘华,白光伟,沈航. 基于消息复用的位置服务隐私保护算法[J]. 微电子学与计算机, 2015, 32(2):148-152.
- [6] 贾金营,张风荔. 位置隐私保护技术综述[J]. 计算机应用研究, 2013, 30(3):641-646.
- [7] 周傲英,杨彬,金澈清,等. 基于位置的服务:架构与进展[J]. 计算机学报, 2011, 34(7):1155-1171.
- [8] Sweeney L. K -anonymity: a model for protecting privacy[J]. International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, 2002, 10(5):557-570.
- [9] 南丽丽,吴涛. 基于信息混淆的社会网络隐私保护机制[J]. 计算机工程与设计, 2011, 32(10):3278-3283.
- [10] Ku W, Zimmermann R, Wang H. Location-based spatial query processing with data sharing in wireless broadcast environments[J]. IEEE Transactions on Mobile Computing, 2008, 7(6):778-791.

作者简介:

杜永文 男,(1974-),博士,副教授.研究方向为嵌入式系统。

朱晓阳(通讯作者) 男,(1986-),硕士研究生.研究方向为位置隐私保护. E-mail:zyxclub@163.com.

练云翔 男,(1987-),硕士研究生.研究方向为无线传感器

网络.